The Design of Data Management System(DMS) at HEPS

Hao Hu, Fazhi Qi, Hongmei Zhang, Haolia Tian, Qi Luo

Computing & Communication System, High Energy Photon Source

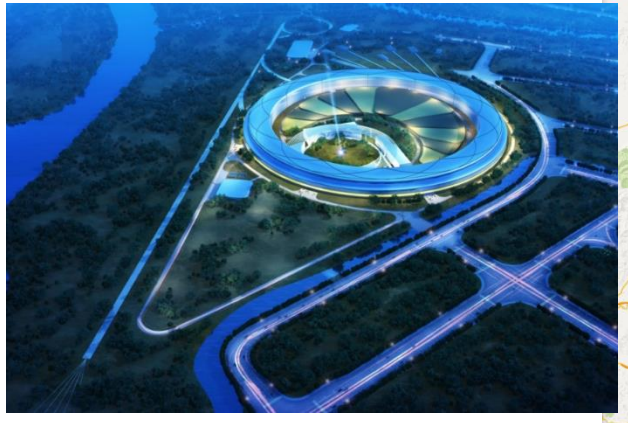
Computing Center, Institute of High Energy Physics

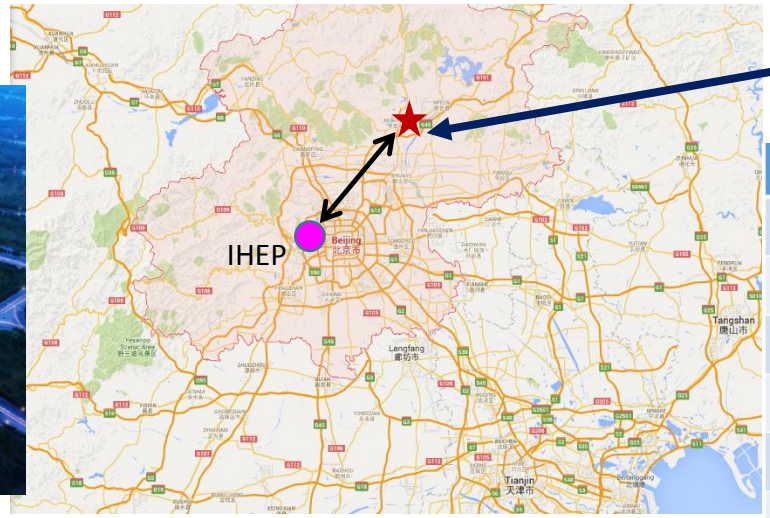
Outline

HEPS： High Energy Photon Source

- New light source in China — High energy, high brightness
- Located in Beijing - about 80KM from IHEP
- Officially approved in Dec. 2017, the construction was started at the end of 2018 and will be completed in 2024
- The whole project will be finished in mid-2025 after commissioning

A new photon science research center at the northern China





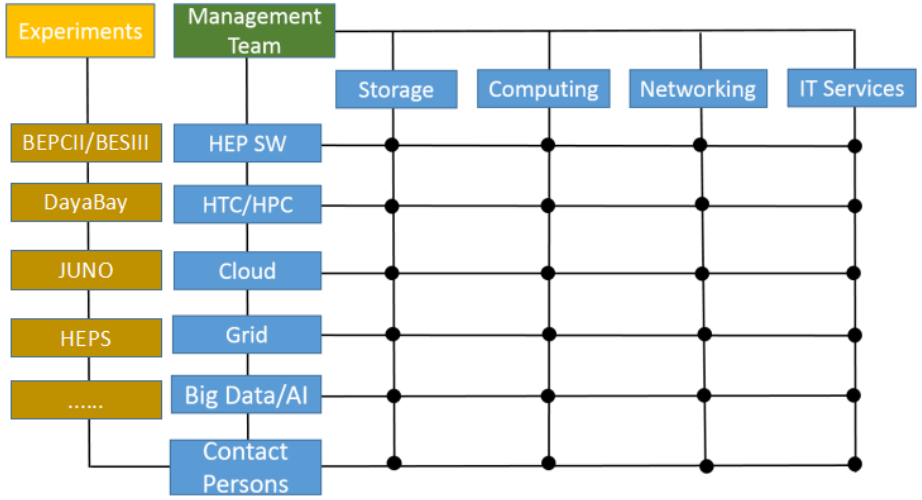
About 80 km from IHEP

·Main parameters	Unit	Value
Beam energy	GeV	6
Circumference	m	1360.4
Emittance	pm·rad	< 60
Brightness	phs/s/mm^2/mrad^2/0.1%BW	>10^{22}
Beam current	mA	200
Injection		Top-up

HEPS CC: the Computing & Communication System for HEPS

- 30+ members
  - All the people are part time for HEPS
  - Most of the members are coming from Computing Center (IHEP CC)
  - 3 from CSNS/Computing and Software group
  - 1 from Beamline
- 9 work groups are set up according to the tasks
  - Infrastructure , Network, Computing & Storage, Scientific Software,

    Data management, Database & Public Service , Monitoring, Security
- Matrix management
  - Across Group Boundaries and Experiments
  - Sharing talents and skills



HEPS Introduction – Beamlines & Data volume

- More than 90 beamlines volume
- Phase I, 14 ID beamlines+1Bending Magnet beamline selected

Beamlines	Burst output(Byte/day)	Average output(Byte/day)
B1 Engineering Materials Beamline	600TB	200TB
B2 Hard X-ray Multi-analytical Nanoprobe (HXMAN) Beamline	500TB	200TB
B3 Structural Dynamics Beamline (SDB)	8TB	3TB
B4 Hard X-ray Coherent Scattering Beamline	10TB	3TB
B5 Hard X-ray High Energy Resolution Spectroscopy Beamline	10TB	1TB
B6 High Pressure Beamline	2TB	1TB
B7 Hard X-Ray Imaging Beamline	1000TB	250TB
B8 X-ray Absorption Spectroscopy Beamline	80TB	10TB
B9 Low-Dimension Structure Probe (LODISP) Beamline	20TB	5TB
BA Biological Macromolecule Microfocus Beamline	35TB	10TB
BB pink SAXS	400TB	50TB
BC High Res. Nanoscale Electronic Structure Spectroscopy Beamline	1TB	0.2TB
BD Tender X-ray beamline	10TB	1TB
BE Transmission X-ray Microscope Beamline	25TB	11.2TB
BF Test beamline	1000TB	60TB
Total average:		**805.4TB/day, 24.16PB/month**

Data management requirement

- Data policy
  - Guidelines for the design and implementation of DMS
- Metadata catalogue
  - Catalogue framework
  - Metadata database
- Metadata ingestor
  - Acquire metadata from DAQ/control system
- Data transfer system
  - Beamline storage → Central storage → Tape
- APIs for interacting with other systems
  - DAQ system, storage system, data analysis system, user service system, proposal system
- Data service
  - Data access, data search, data download

HEPS Data Policy

The ownership, curation, archiving and access to scientific data and metadata

- Data classification: raw data, processed data, calibration data, result data
- Recommend providing at least 3 months disk storage and permanent tape archive (depends on final funding)
- Provide permanent storage for raw data, calibration data and result data
- Provide temporary storage for processed data
- Each dataset will have a unique persistent identifier(CSTR/PID21/doi)
- Experimental teams have sole access to the data during the embargo period. After the embargo, the data will be released with open access to any registered users of the HEPS data portal.

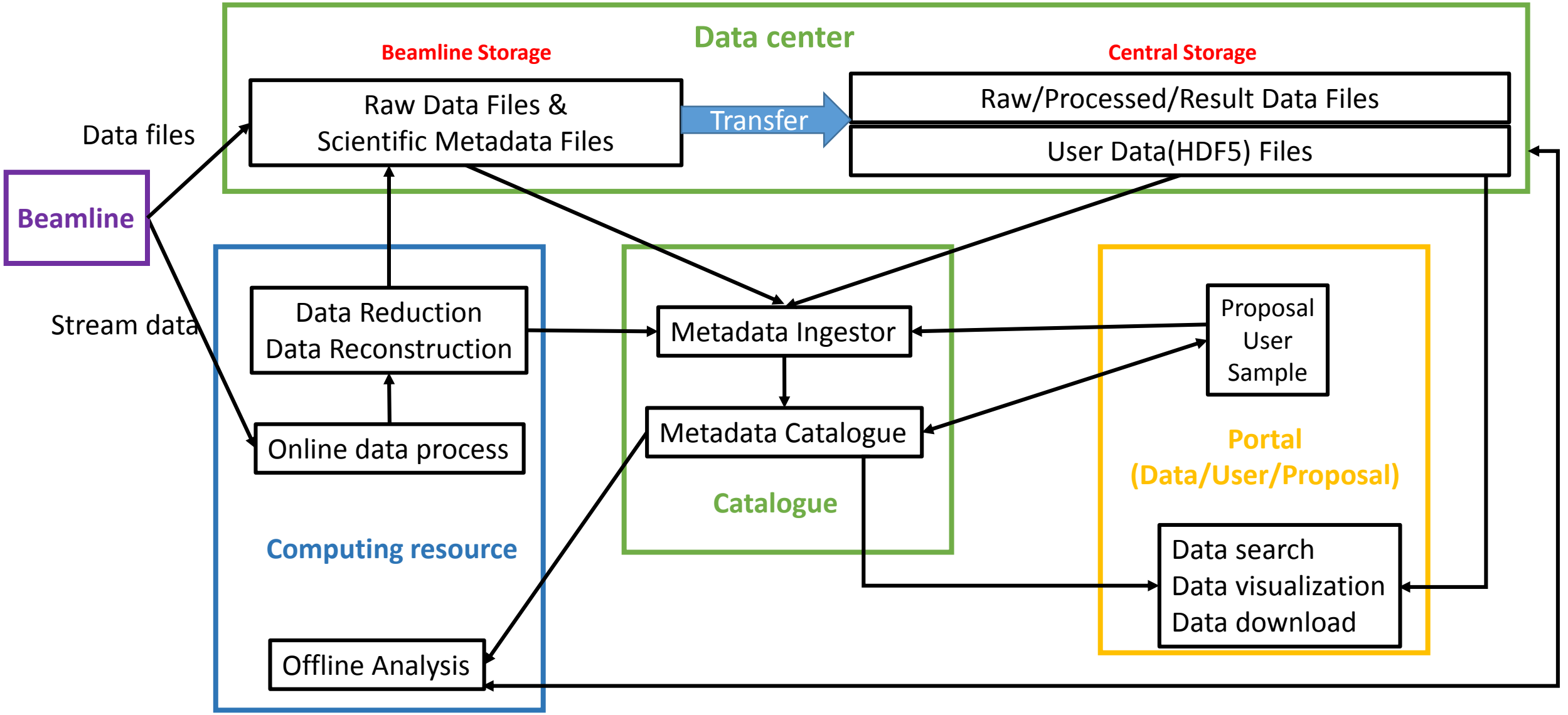
A draft version of ***The Data Policy for HEPS*** is finished, which will be discussed and approved by the HEPS council.

Reference:

http://pan-data.eu/sites/pan-data.eu/files/PaN-data-D2-1.pdf

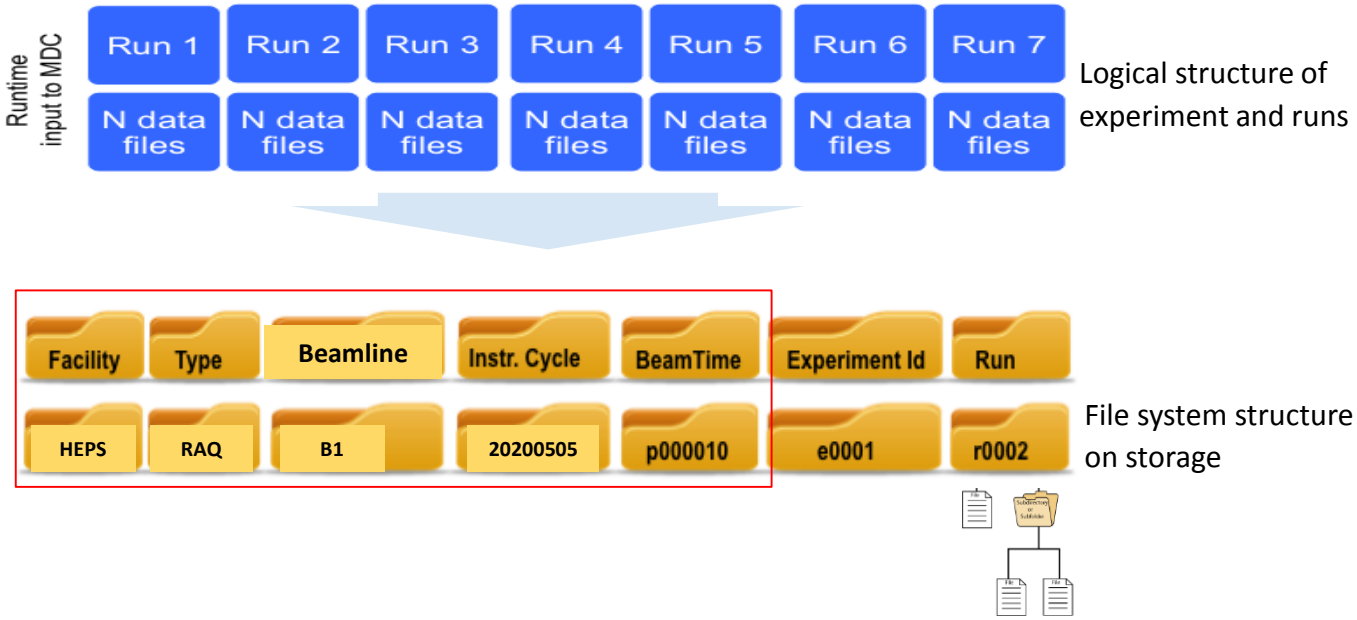
https://in.xfel.eu/upex/docs/upex-scientific-data-policy.pdf

Data Flow


Data center
Beamline Storage
Central Storage
Raw Data Files &
Scientific Metadata Files
Transfer
Raw/Processed/Result Data Files
User Data(HDF5) Files
Data files
Beamline
Stream data
Data Reduction
Data Reconstruction
Online data process
Computing resource
Metadata Ingestor
Metadata Catalogue
Catalogue
Proposal
User
Sample
Portal
(Data/User/Proposal)
Data search
Data visualization
Data download
Offline Analysis

Predefined directory structure

- Beamline file system: Lustre/NFS
- Central file system: Lustre
- Directory structure of data files on storage



Logical structure of experiment and runs

File system structure on storage

```
|--hepsfs
    |-- <Beamline>
        |-- <YEAR>
            |-- <BeamtimeID>
                |-- raw
                    |--<Sample>
                        |--<ScanID>
                            |--data files…
                |-- processed
                |-- scratch
                |-- share
```

Beamline file system & Central file system

Reference: *A. Rothkirch, Handling the Data Deluge in Photon Science Experiments: Current Status and Future Plans at DESY, talk given at the 12th NOBUGS Conference, Brookhaven National Laboratory, October 22–26, 2018.*

Metadata

- **Administrative metadata**

  proposal Info, experiment group, measurement technique, beamline; dataset Info(PID, format, size, path, creation time);

  data files (file size, format, storage path, time, checksum);

  analysis software, update time…

  From:

  Proposal system, User service system, File transfer tools, Analysis software

  For: catalogue

  To: catalogue database

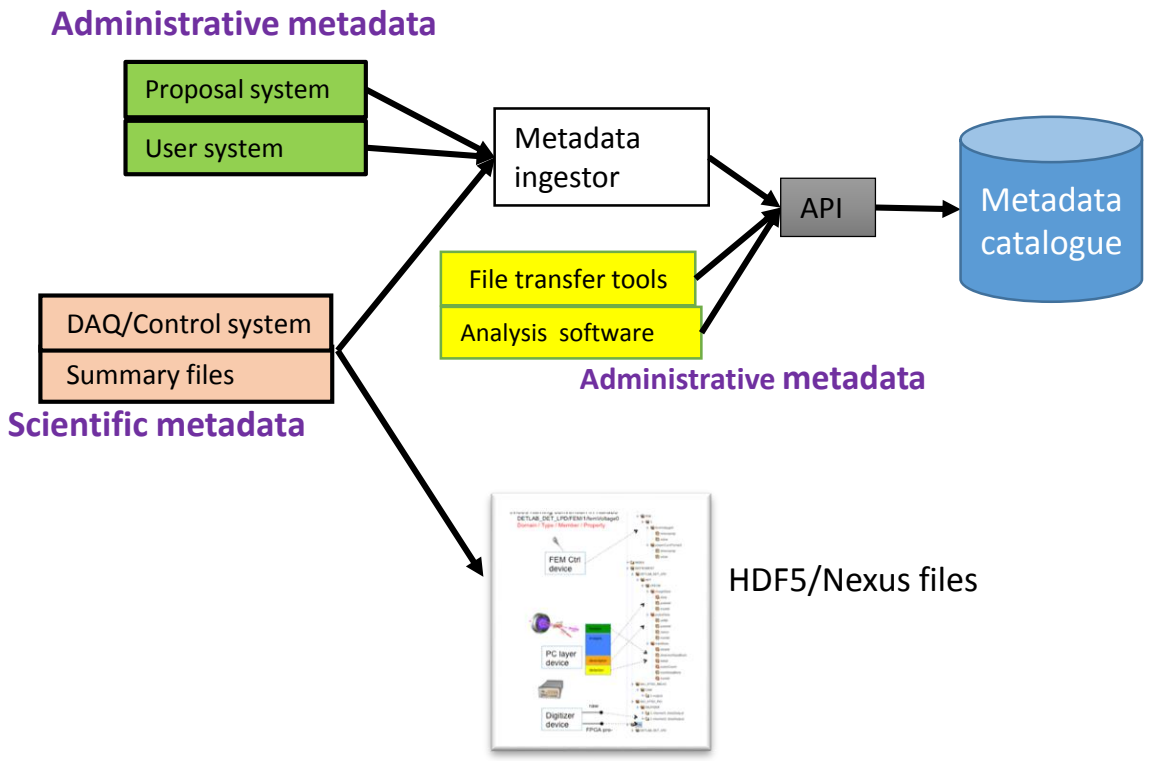
- **Scientific metadata**

  Sample info, Experiment environment parameters…

  From: DAQ/Control system

  For: analysis & catalogue

  To: files(HDF5/Nexus) , catalogue database



Q: What kind of metadata are needed cataloging?

A: Those metadata necessary and significant for data searching and sharing!

Metadata catalogue framework--SciCat

SciCat： open source, developed by PSI and ESS (and MAX IV) to create a data catalog management system

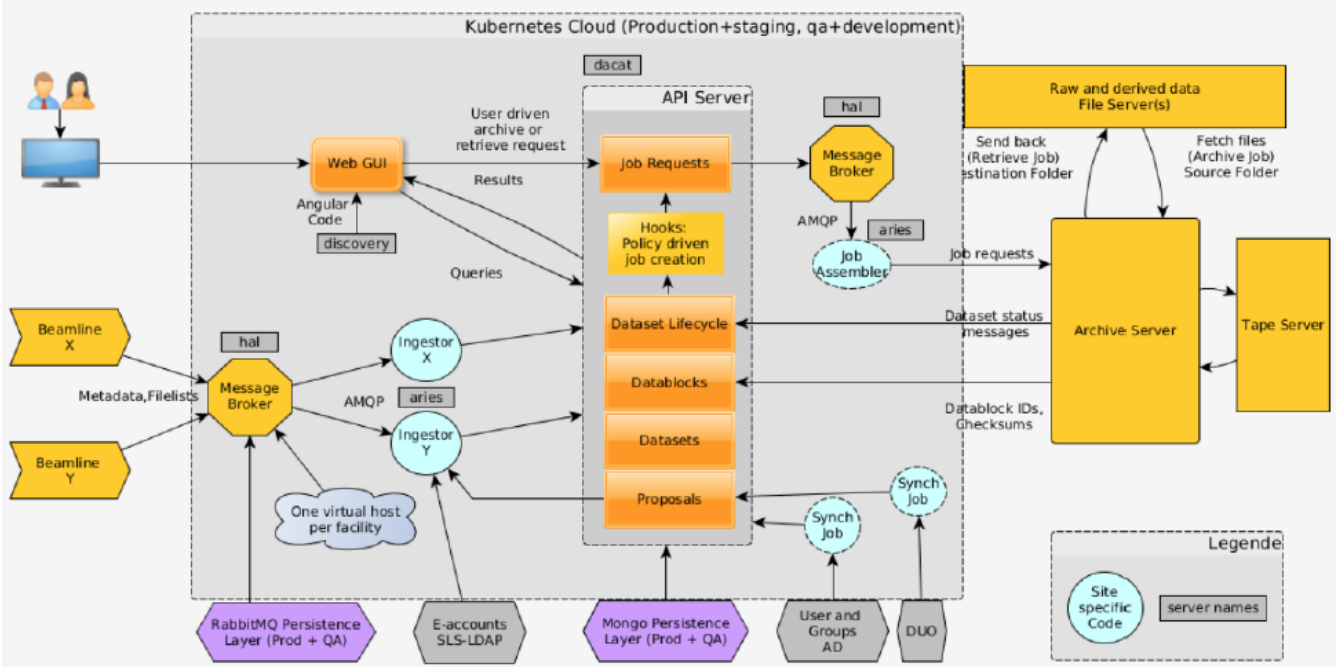
- Message queue: RabbitMQ
- LoopBack
  - metadata models are defined in JSON format
  - RESTful APIs are exposed via a NodeJS web server
- NoSQL database: MongoDB
- Web front end: Angular8

HEPS:

Kafka as message queue

use LoopBack to create REST APIs

MongoDB as the metadata database



https://scicatproject.github.io/

Metadata acquisition approaches

Depends on how metadata are provided

- Interfaces are provided for control system to write metadata to DMS(recommended)
- Metadata-ingestor plugins are designed to collect metadata from nexus/txt/HDF5 files

Metadata acquisition approach（1）

- Suitable for self-designed detector, having ability to program to control software(Epics/Tango）
- Interfaces with control system are provided
- With which control system can send metadata (JSON-based) to kafka broker when a dataset is produced

README.md

Example from github

Area detector Kafka interface

This repository contains two separate projects which facilitates the transmission of data between an EPICS IOC and an Apache Kafka broker. The two projects are:

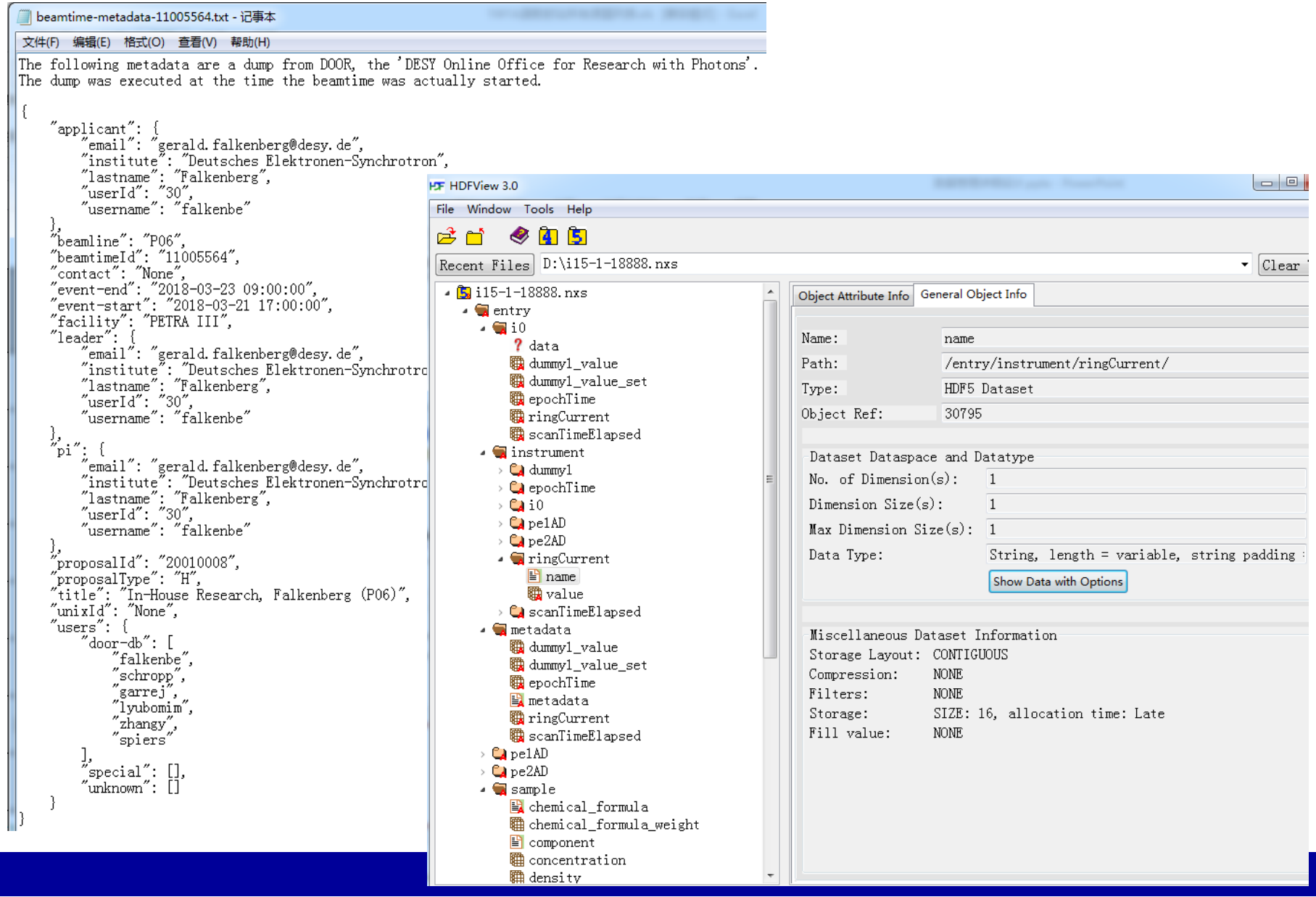
- An EPICS areaDetector driver which acts as a Kafka consumer and makes NDArray data received from the broker available to the IOC.
- An EPICS areaDetector plugin which connects to an areaDetector and serializes NDArray data it receives and sends it to a Kafka broker.

Apache Kafka is an open-source platform for handling streaming data one or more data brokers in order to maximize throughput and reliability. More information on Apache Kafka can be found at the website of that project.

For serializing and de-serializing the areaDetector (NDArray) data, Google FlatBuffers is used. Serializing data using FlatBuffers is fast with a relatively small memory overhead while being easier to use than C-structs.

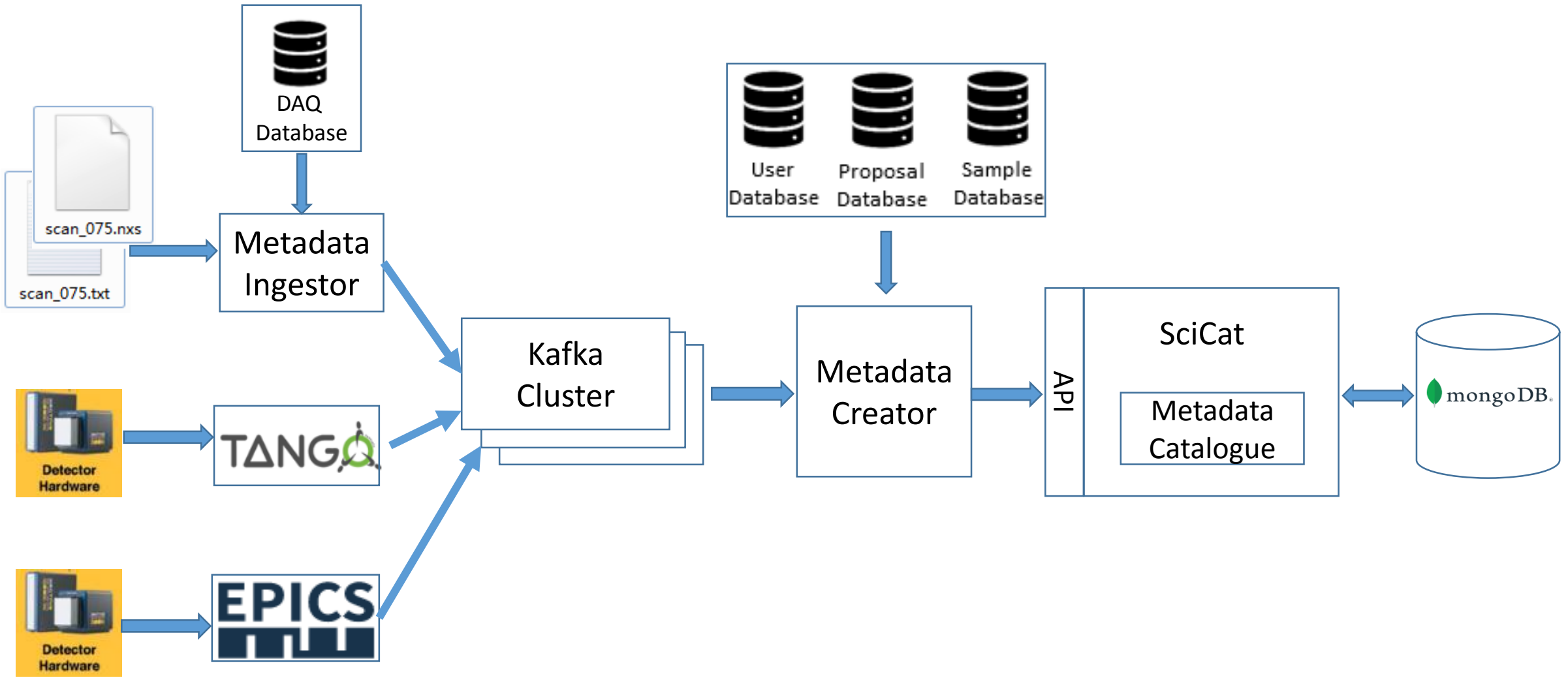
```
{ topic: 'bm01_detector1_topic',
  messages: [{
            "proposalID":"p0001",
            "sampleID":"s0001",
            "userAccount":"wch@ihep.ac.cn",
            "creationLocation":"bm01_detector1"
            "file_name":"bm01_1_38883",
            "dataFormat":"hdf5",
            "scientificMetadata":{
                        elog_id:"234",
                        optical_coupling:"dryfit"},
            "description":"dataset description"
            }]
 }
```

Metadata acquisition approach（2）



- Suitable for those that metadata are collected and saved to files (txt/nxs)
- Metadata-ingestor runs in flexible plugin mode.
- In addition to user and experiment information, the plug-in also records information such as file size, checksums, processing time, and file replicas.

Metadata acquisition— for cataloging



Data format standardization

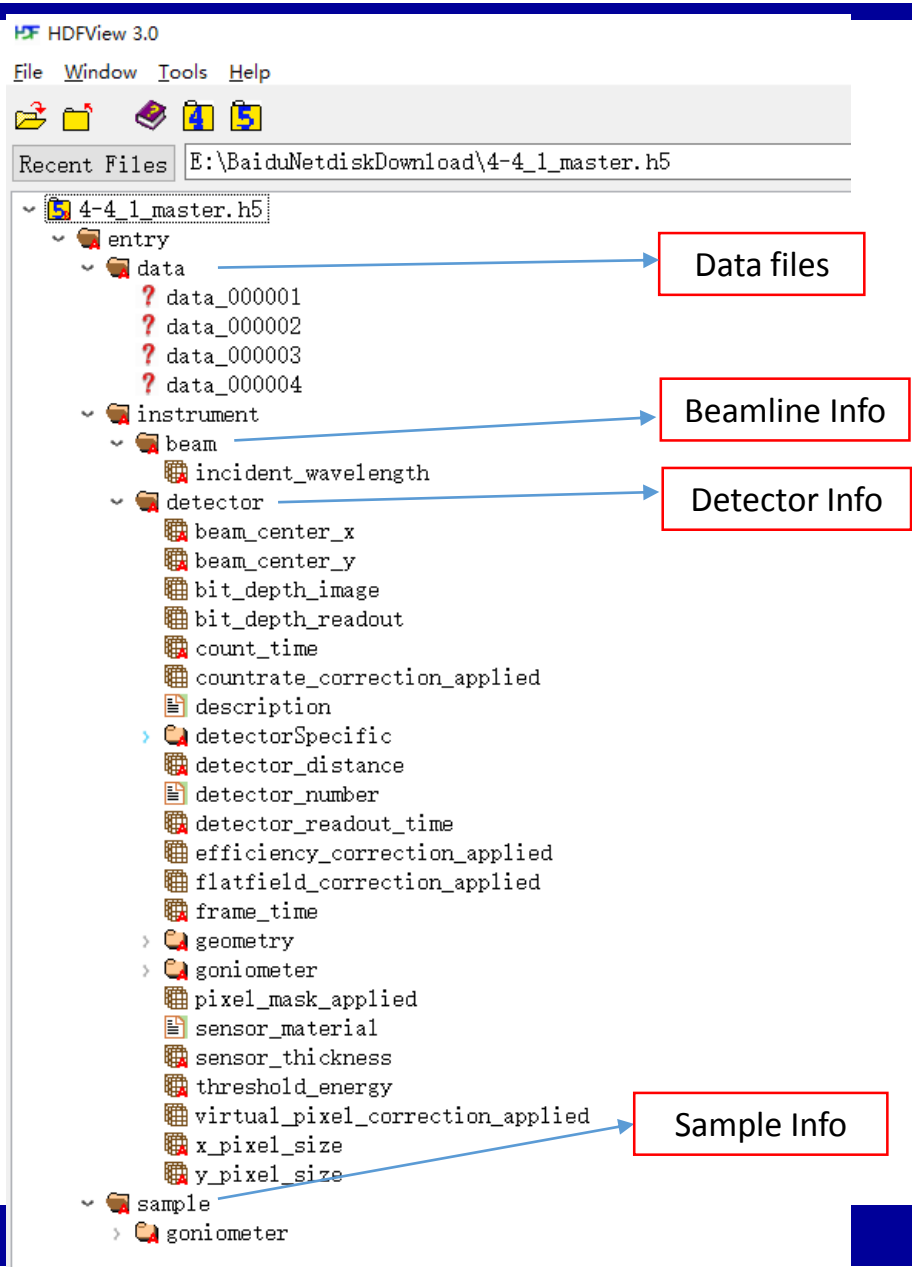
- Data formats used by X-ray experiments:

  txt/tif/binary/dat/jpg/hdf5/…

- **HDF5** is chosen as the standard data format
- HDF5 file schema follows NeXus conventions
  - NeXus is a common data format for neutron, x-ray, and muon science
  - https://www.nexusformat.org/
  - NeXus add meaning to HDF5
- Data + beamline + detector + Sample

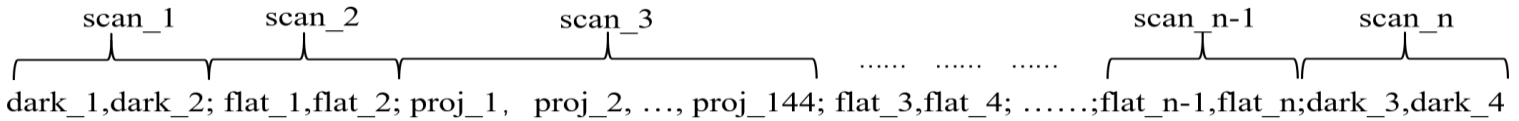
HDF5 is entirely agnostic about what data is stored!

Nexus is a set of rules how data should be organized in an HDF5 file.



HDF5 file format design for B7(Hard X-Ray Imaging Beamline)

CT data consists of three types:
dark—images without X-ray
flat– images without sample
Projection– images with X-ray and sample

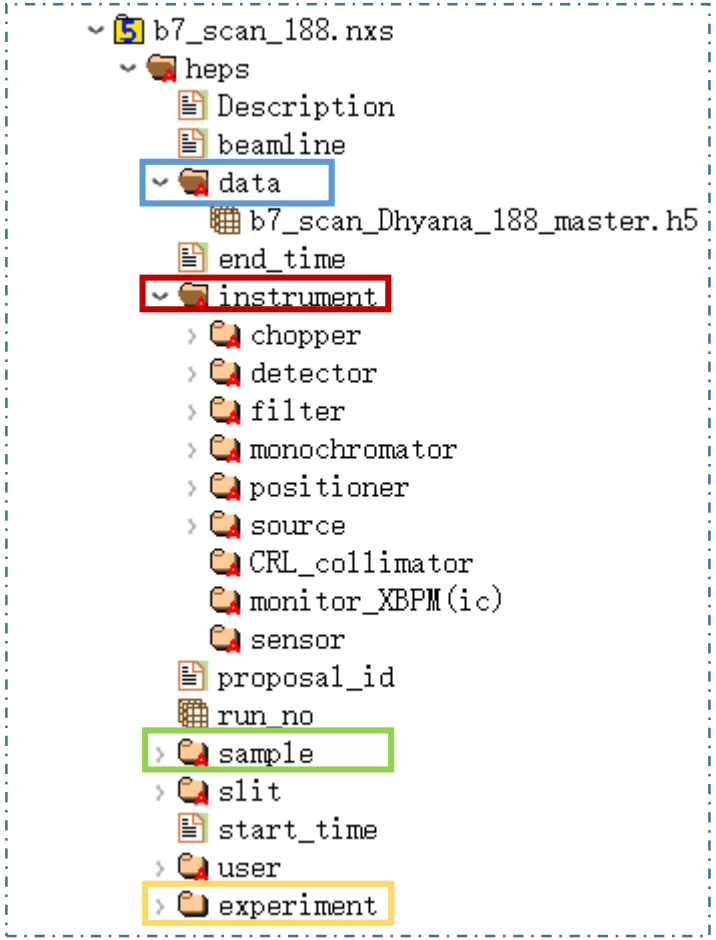


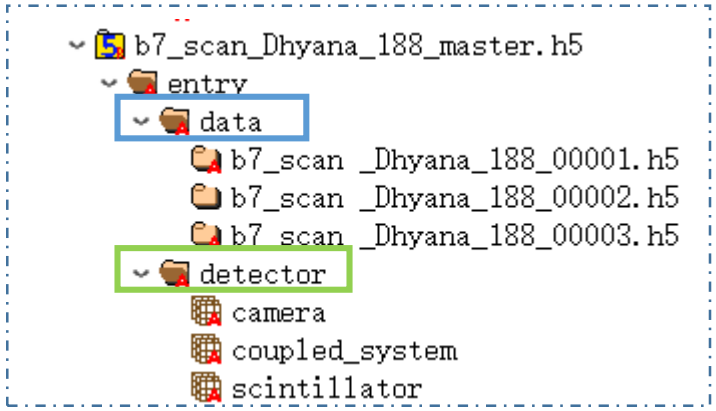
Demo: a set of CT data files

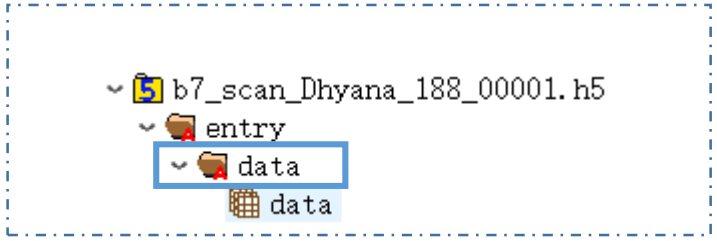
b7_scan_*188*.nxs

b7_scan_*Dhyana_188*_master.h5

b7_scan _*Dhyana_188*_00001.h5
b7_scan _*Dhyana_188*_00002.h5
…
…
b7_scan _*Dhyana_188*_02000.h5







Need more discussion and verification

DMS testbed at 1W1A of BSRF

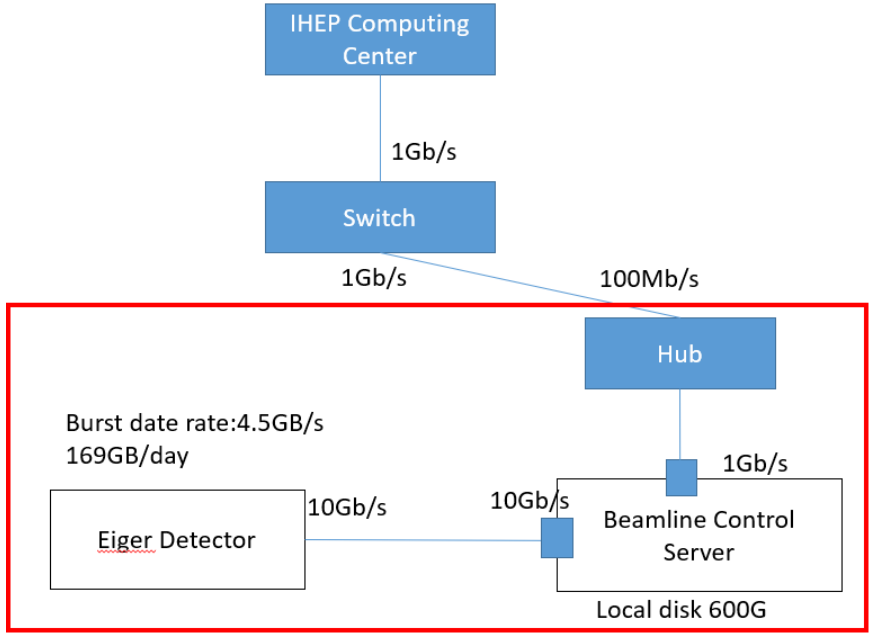
BSRF: Beijing synchrotron radiation facility, since 1991, IHEP

provides the technology R&D and test platforms for HEPS

1W1A: a diffuse X-ray scattering beamline of BSRF

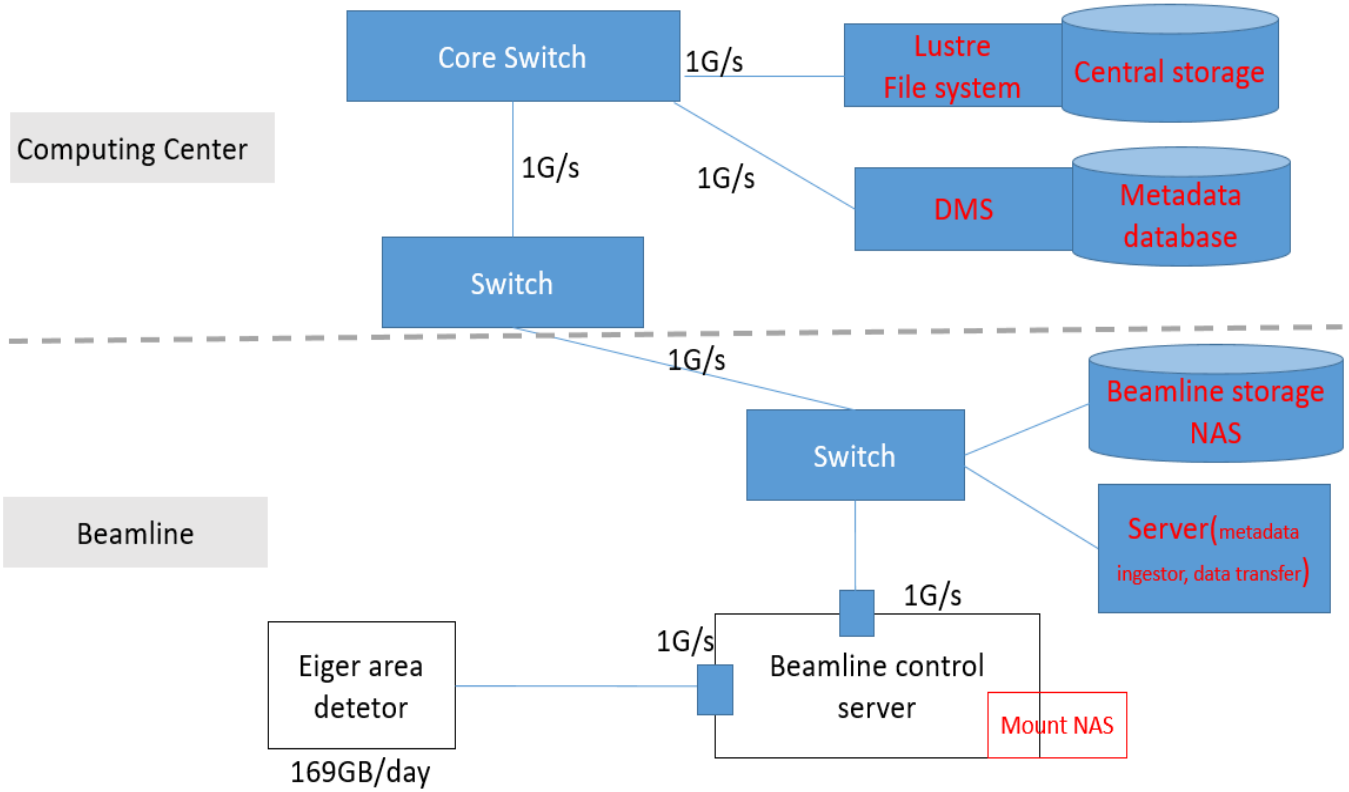
Eiger area detector, 170GB raw data per day

Task： verify DMS and the whole process of data acquisition, data transfer, data storage and data access



Progress of testbed at 1W1A

1. Network bandwidth is upgraded from 100Mb/s to 1Gb/s
2. Beamline storage: **2TB** NAS, Dell EMC NX3240, NFS file system
3. Central storage: **80TB** disk array, Lustre file system, located at Computing Center
4. A server located at the beamline, responsible for
    - metadata ingesting
    - data transferring
5. DMS API server and database server, located at Computing Center
6. Data service web portal
    - authorized users and administrators
    - access data, download data files



Web front end

查看数据

文件名	创建时间	路径	大小(K)	下载
TOMO_183_1237.h5	2020-12-14 11:40:29	/hepsfs/beamline-test/2020/beamtimeID01/raw/TOMO_183_1237.h5	4002.0	下载
TOMO_183_344.h5	2020-12-14 11:40:29	/hepsfs/beamline-test/2020/beamtimeID01/raw/TOMO_183_344.h5	4002.0	下载
TOMO_183_489.h5	2020-12-14 11:40:29	/hepsfs/beamline-test/2020/beamtimeID01/raw/TOMO_183_489.h5	4002.0	下载
TOMO_183_960.h5	2020-12-14 11:40:29	/hepsfs/beamline-test/2020/beamtimeID01/raw/TOMO_183_960.h5	4002.0	下载
TOMO_183_644.h5	2020-12-14 11:40:29	/hepsfs/beamline-test/2020/beamtimeID01/raw/TOMO_183_644.h5	4002.0	下载
TOMO_183_1011.h5	2020-12-14 11:40:29	/hepsfs/beamline-test/2020/beamtimeID01/raw/TOMO_183_1011.h5	4002.0	下载
TOMO_183_825.h5	2020-12-14 11:40:29	/hepsfs/beamline-test/2020/beamtimeID01/raw/TOMO_183_825.h5	4002.0	下载
TOMO_183_973.h5	2020-12-14 11:40:29	/hepsfs/beamline-test/2020/beamtimeID01/raw/TOMO_183_973.h5	4002.0	下载
TOMO_183_844.h5	2020-12-14 11:40:29	/hepsfs/beamline-test/2020/beamtimeID01/raw/TOMO_183_844.h5	4002.0	下载
TOMO_183_1199.h5	2020-12-14 11:40:29	/hepsfs/beamline-test/2020/beamtimeID01/raw/TOMO_183_1199.h5	4002.0	下载
TOMO_183_628.h5	2020-12-14 11:40:29	/hepsfs/beamline-test/2020/beamtimeID01/raw/TOMO_183_628.h5	4002.0	下载
TOMO_183_725.h5	2020-12-14 11:40:29	/hepsfs/beamline-test/2020/beamtimeID01/raw/TOMO_183_725.h5	4002.0	下载
TOMO_183_145.h5	2020-12-14 11:40:29	/hepsfs/beamline-test/2020/beamtimeID01/raw/TOMO_183_145.h5	4002.0	下载
TOMO_183_1030.h5	2020-12-14 11:40:29	/hepsfs/beamline-test/2020/beamtimeID01/raw/TOMO_183_1030.h5	4002.0	下载
TOMO_183_780.h5	2020-12-14 11:40:29	/hepsfs/beamline-test/2020/beamtimeID01/raw/TOMO_183_780.h5	4002.0	下载
TOMO_183_876.h5	2020-12-14 11:40:29	/hepsfs/beamline-test/2020/beamtimeID01/raw/TOMO_183_876.h5	4002.0	下载
TOMO_183_513.h5	2020-12-14 11:40:29	/hepsfs/beamline-test/2020/beamtimeID01/raw/TOMO_183_513.h5	4002.0	下载
TOMO_183_68.h5	2020-12-14 11:40:29	/hepsfs/beamline-test/2020/beamtimeID01/raw/TOMO_183_68.h5	4002.0	下载

Summary and Plan

- The Data Management System has passed FDR
- Fundamental functions are implemented on the testbed
- Optimize the high availability and reliability of the system
- Specify the HDF5 data format schema for each experiment technique (cooperate with beamline scientists)
- Testbed at beamline 3W1 of BSRF is under way
  - Integrate with the data analysis framework and computing system
  - verify the design and performance of the IT infrastructure
  - the work on testbeds will directly apply to HEPS
- We benefit a lot from the papers and reports of other Photon Source Facilities
- We hope to have chance to cooperate with other Facilities, like PSI/MAXIV/DESY etc.

Thanks for your attention!

Any comments or suggestions?